A Summary of the Method Used to Establish Cut Score Recommendations for the TCAP ACH and MAAS Assessments

2010 LEAD Conference

Erika Hall, PhD

PEARSON

Goal for this session:

- Provide a general introduction to Standard Setting
- Provide the context for the TN ACH and MAAS Standard Setting
  - Why necessary?
- Describe the procedure used to establish cut score recommendations
  - Bookmark

Basic Vocabulary

- **Content Standards:** the content and skills that students are expected to know and be able to do.
- **Achievement Levels** (Performance Levels, Performance Categories): Labels for levels of student achievement (e.g., basic, proficient and advanced).
- **Achievement Level Descriptors** (ALDs): Descriptions of the competencies associated with each level of achievement.
- **Cut Scores** (Standards): Scores on an assessment that separate one level of achievement from another

Basic Vocabulary (Continued)

- **Panelists (Judges/Raters):** Those who participate in the committee-based component of the standard setting process (stakeholders, educators, professionals – must understand the content assessed).
- **Feedback Data:** Data provided to panelists to help them analyze the validity and reasonableness of the standards they are recommending (e.g., Median/Mean cutscore ratings, table agreement, etc.)
- **Impact Data (Normative Feedback):** Data that summarize the consequences of a proposed set of cutscores. (How many students will be classified below, at or above proficient?)

Basic Vocabulary (Continued)

- **External Reference Data** - Performance data associated with tests having similar goals and student expectations as the test for which standards are being set (e.g., NAEP, ACT, EXPLORE, PLAN)
  - Important to consider because it is relevant and useful
  - Not intended to dictate results or be considered a target
  - Serves as a reality check and/or point of discussion

What is Standard Setting?

- A judgmental, value-based *process* which has a variety of steps and includes relevant stakeholders throughout. Steps in this process include:
  1. Identifying the relevant knowledge and skills to be taught and assessed at each grade/content area to support the goals of the state
  2. Defining the expectations associated with each Achievement Level
  3. Convening a committee of educators to provide content-based *recommendations* for cut scores at each grade
  4. Convening policy makers and other stakeholder groups to review the impact associated with the recommended cut scores
  5. Review of the results from steps 1-4 by the TN Education Commissioner who makes recommendation to State Board
  6. Review and approval of cut score recommendations by the State Board of Education

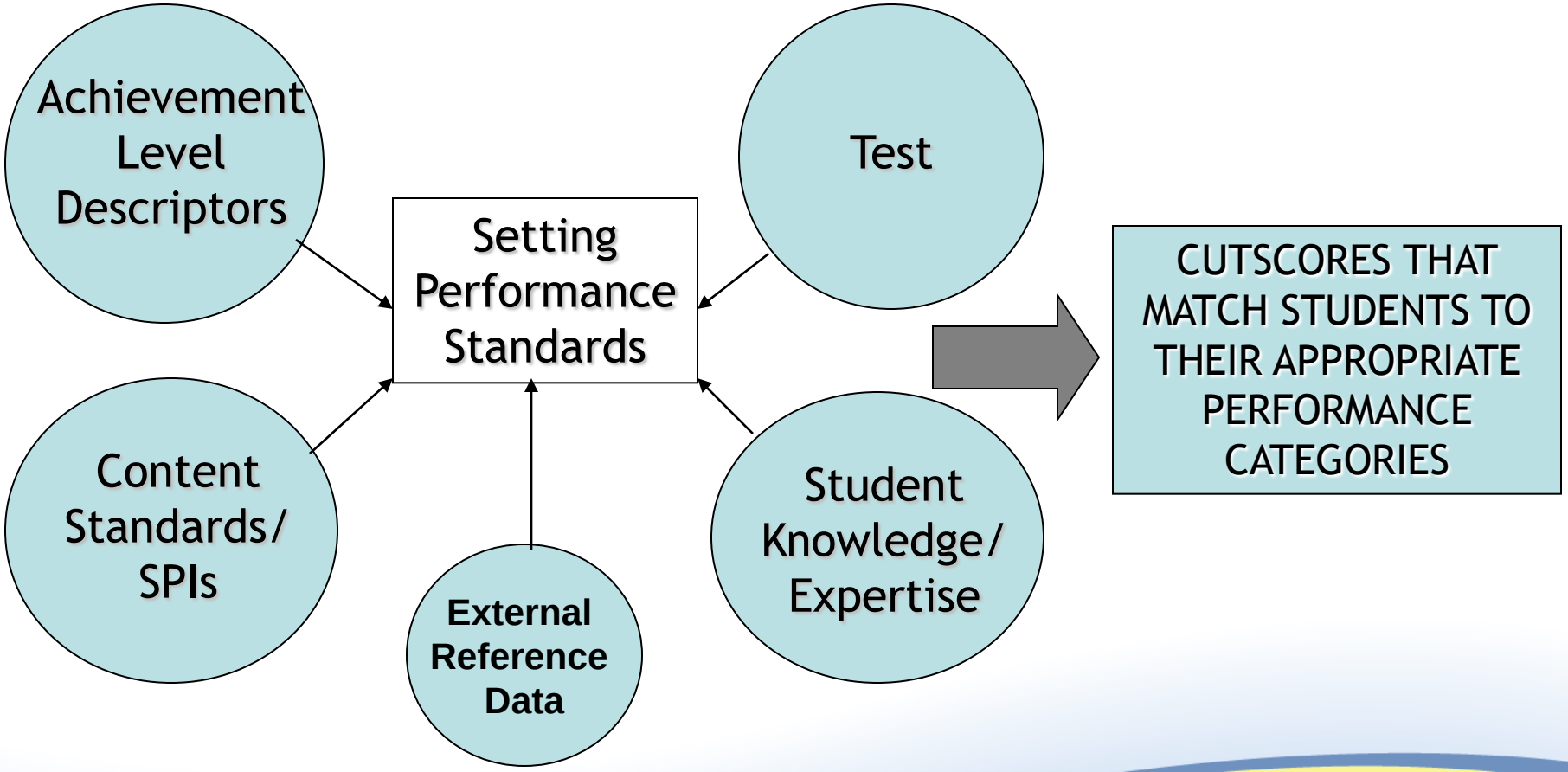
Goal of Standard Setting Meeting

- Use a well-defined, legally defensible procedure to obtain cut score *recommendations* from those in the best position to make them (those who know the range of abilities in the test taking population with respect to the knowledge and skills assessed).

- To quantify student performance expectations on the reportable scale.
  - Operationalize the Achievement Level Descriptors

- Obtain evidence for the validity of (some) inferences and decisions made in consideration of defined achievement levels.

Setting Performance Standards (Cut Scores)



When is a Standard Setting Necessary?

- New Assessment (MAAS)
- Existing Assessment (ACH)
  - Curriculum Updates
  - Changes to Test Design or Content
  - New Federal Requirements
  - Increased Expectations for Performance

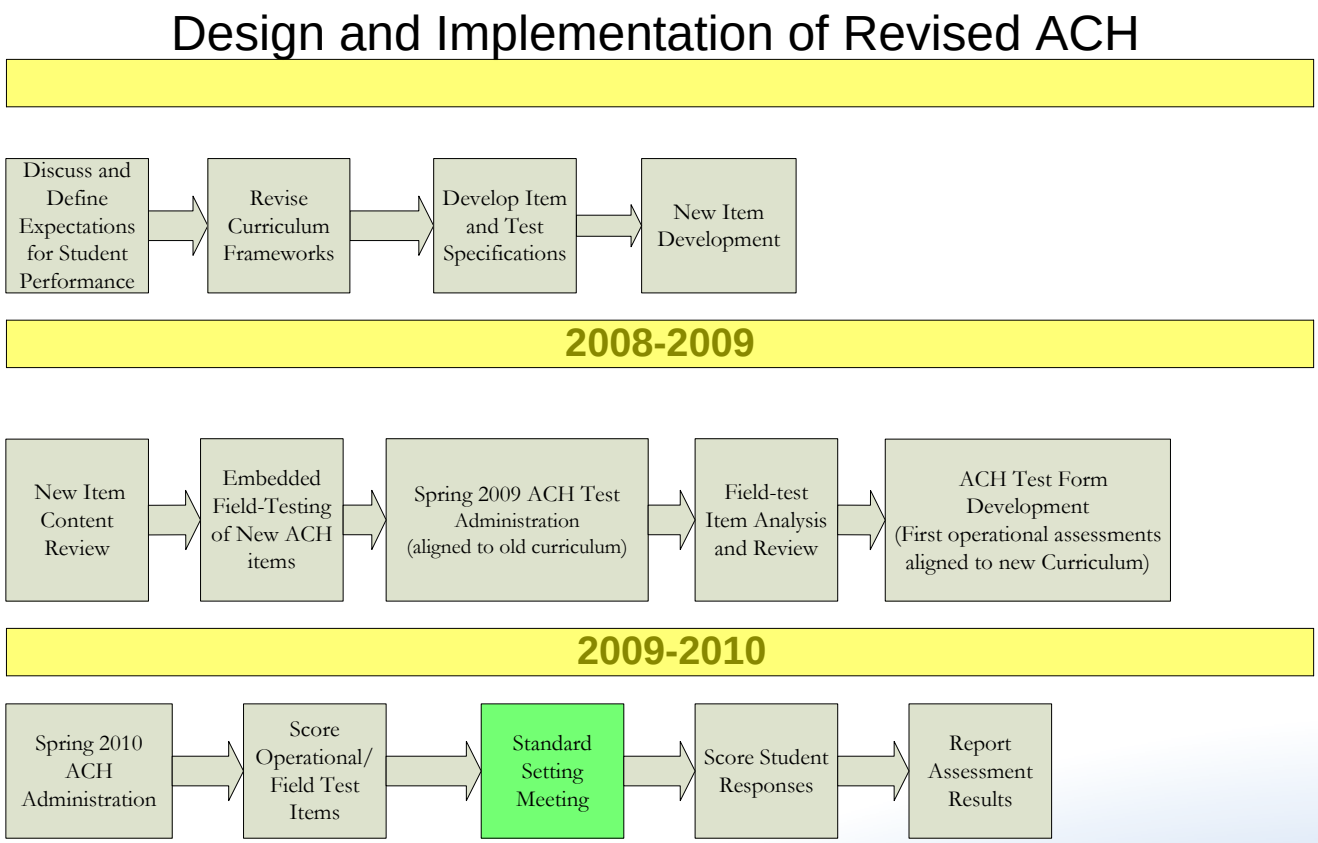
Context for ACH Standard Setting

- In light of TN goal to teach and assess skills associated with college and career readiness, new curriculum frameworks were established in Mathematics, Reading Language Arts and Science
- This necessitated the development of new ACH assessments for these content areas
  - Aligned with new content standard and SPIs
  - Representing the appropriate degree of cognitive complexity
- Previously established cuts cores no longer apply
  - New test measuring something "different" old test
  - Four achievement levels rather than three
  - What it means to be "Proficient" has changed in light of the goals of the new ACH assessments

Context for MAAS Standard Setting

- MAAS – Modified Academic Achievement Standards Assessment
  - Intended for a small group of students with an IEP for whom:
    - the alternate assessment is too easy
    - appropriate instruction is unlikely to result in grade-level proficiency
  - Goal: Provide for more information about what these students know and can do relative to the grade-level content standards
- First MAAS operational assessment in Spring 2010

Where in Process does Standard Setting Occur?



Design of the Meetings

- 3 committees for each content area (Science, Math and Reading Language Arts, Social Studies) defined by grade bands (3/4, 5/6, 7/8)
  - 12 MAAS committees – 164 participants
  - 9 ACH committees (no Social Studies) – 140 participants

MAAS/ACH Standard Setting Schedule

Monday, June 21st	Tuesday, June 22nd	Wednesday June 23rd		Thursday, June 24th	Friday, June 25th	
8:30-5:30	8:00-5:00	a.m.	p.m.	8:00-5:00	a.m.	p.m.
MAAS (SS)	MAAS (SS)	MAAS (VA)				
		ACH (SS)	ACH (SS)	ACH (SS)	ACH (VA)	

General Flow of Standard Setting Meeting

First - General overview that describes the purpose of standard setting and provides background on the tests under consideration. (1-2 hours).

Second - Panelists are broken down into grade-subject specific groups to work on standard setting tasks. They:

- Discuss and clarify the Achievement Levels under consideration (2 hours)
- Receive training/practice on the standard setting method (2 hours)
- Apply the standard setting methodology to recommend cutscores in an iterative fashion (3 rounds, over 2 days)

Third – Vertical Articulation

- Review recommendations over grades (1/2 day)

Steps in the TN Bookmark Process

1. Review of the Achievement Level Descriptors
   - General Achievement Level Descriptors
   - Specific Achievement Level Descriptors
     - Define expectations specific to a given grade/course

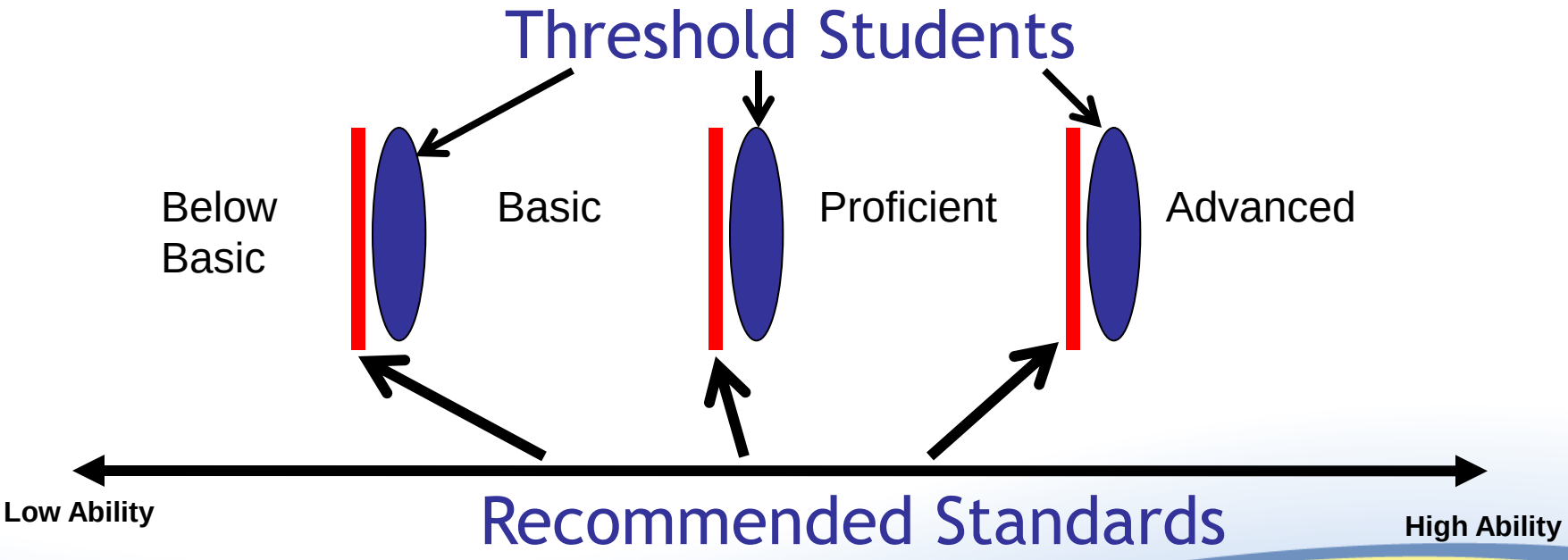
Goal of review is to understand:

- The knowledge, skills and capabilities that define a typical student at each level
- The key factors that distinguish students at adjacent levels
- The range of student abilities represented at each level given the ALDs

Steps in the Bookmark Process (cont.)

2. Define the Threshold-Student for each level.

- Borderline or minimally qualified student in terms of performance

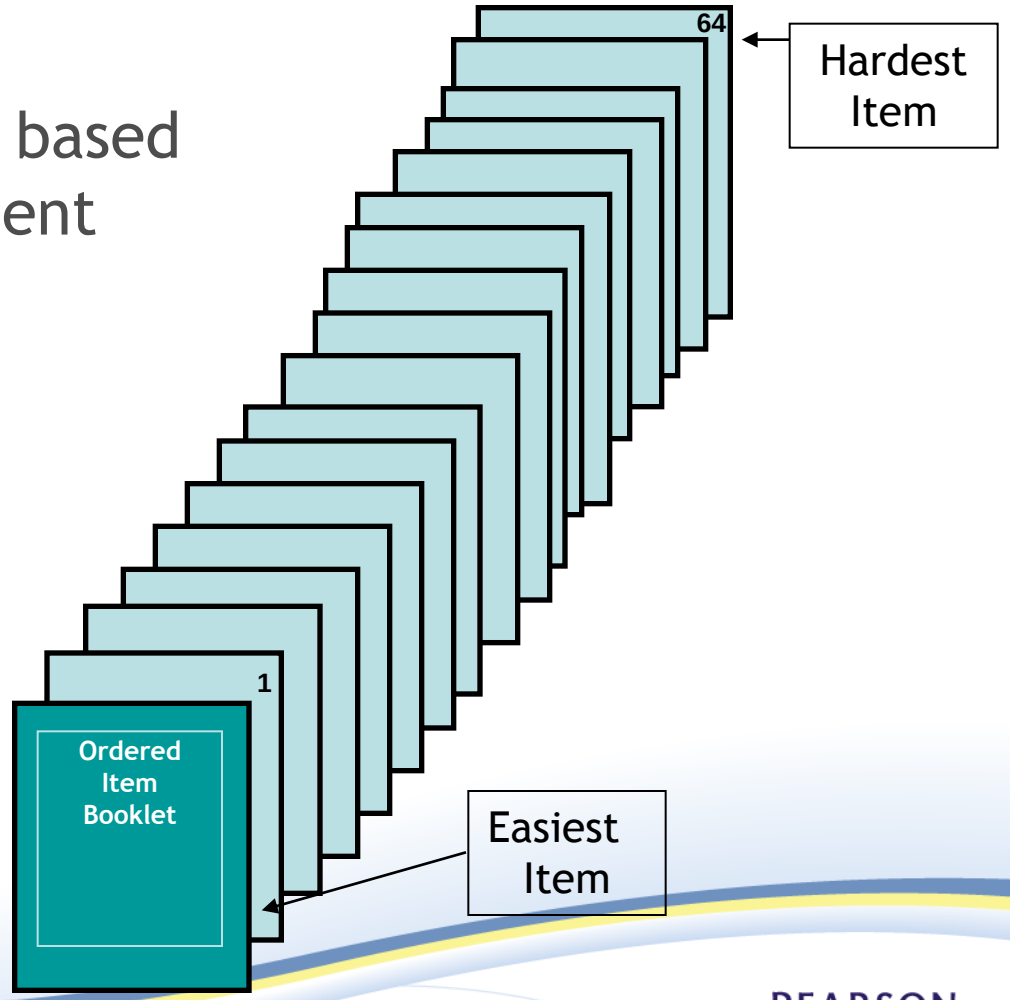


Steps in the Bookmark Process (cont.)

3. Training on the Bookmark recommendation procedure
   - Reiterate purpose
   - Introduce materials
   - Explain the recommendation process
   - Practice implementing the procedure
   - Group discussion on process
   - Answer all questions

Ordered Item Book

- 2010 operational test items
- Items ordered based on actual student performance
- 64-75 items
- Represents a continuum of skills and abilities.



Sample OIB Page

Bookmark Process

Goal:

- Identify location in OIB that best represents each cut score – or the transition from one level to the next.
  - The place in the OIB that accurately divides the items into those that **all students** at a given level are likely to answer correctly from those they are not likely to answer correctly
    - Likely defined as 2/3 of the time or greater

Standard Setting - Process

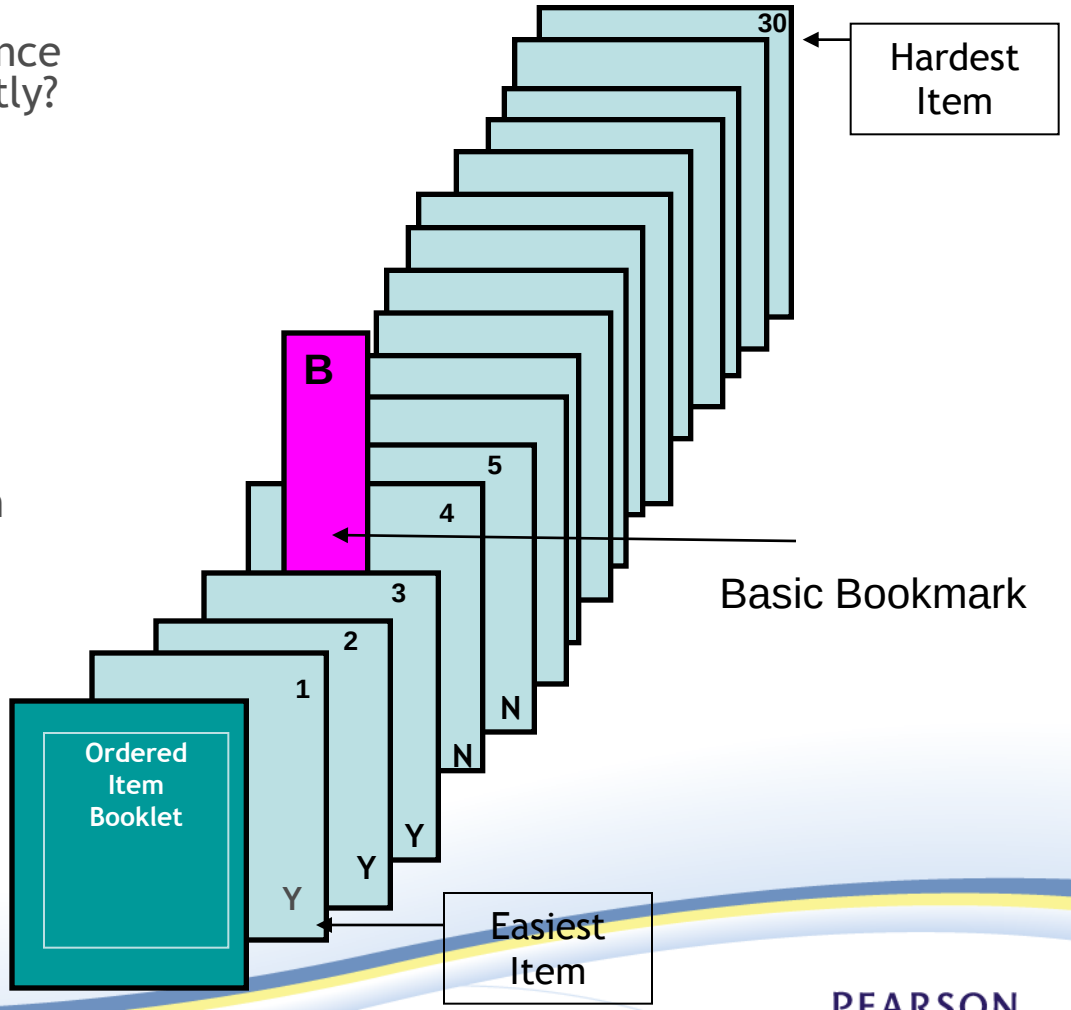
- Visualize a typical student at the threshold of a given level
- Recall the characteristics of this typical threshold student (discussed in large group)
- Identify the last item in the OIB you believe a typical threshold student is likely to answer correctly (i.e., has 2/3 chance or greater)
- Place the bookmark after this item.
- Write down the page number in front of your maker on the Bookmark Recommendation Sheet

Determining the Bookmark Location (cont.)

- Think about typical threshold Basic student.
- Would he/she have a 2/3 chance of answering this item correctly?

1. Yes
2. Yes
3. Yes
4. No
5. No

- Basic bookmark goes between Pages 3 and 4.
- Bookmark page number is 3.



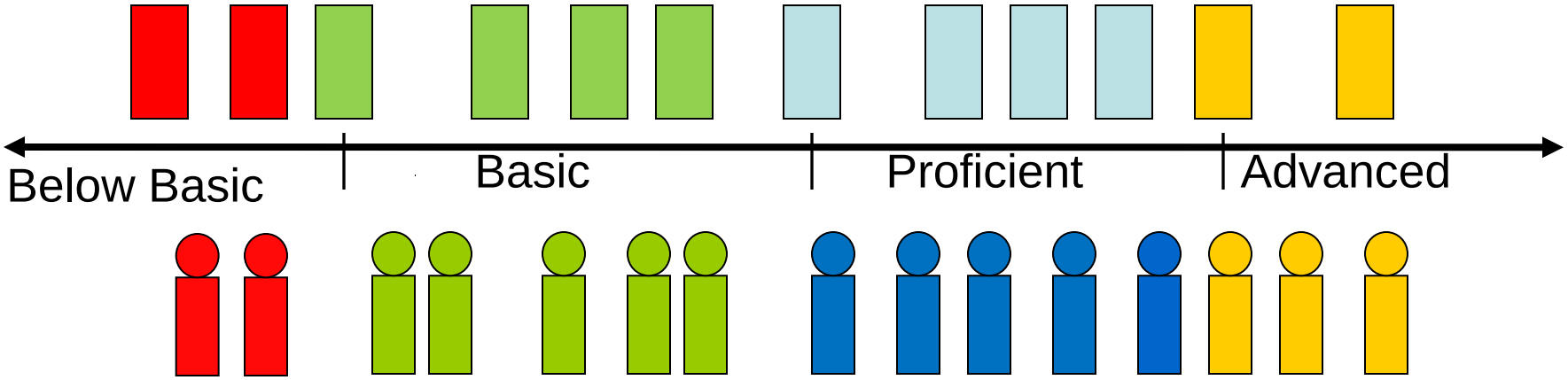
Theory vs. Practice

Theory	
Page	Answer
19	Y
20	Y
21	Y
22	Y
23	N
24	N
25	N

Practice	
Page	Answer
19	Y
20	Y
21	N
22	N
23	N
24	Y
25	N

Rationale for Bookmark Process

OIB – Items are ordered from least difficult to most difficult



Students ordered by knowledge and skills

Statistical model allows item difficulty and student ability to be placed on the same scale.

Steps in the Bookmark Process (cont.)

4. Review of the Ordered Item Book
   - Think about the knowledge and skills required to answer each item correctly
5. Complete Readiness Survey
6. Make first round of recommendations using Bookmark method

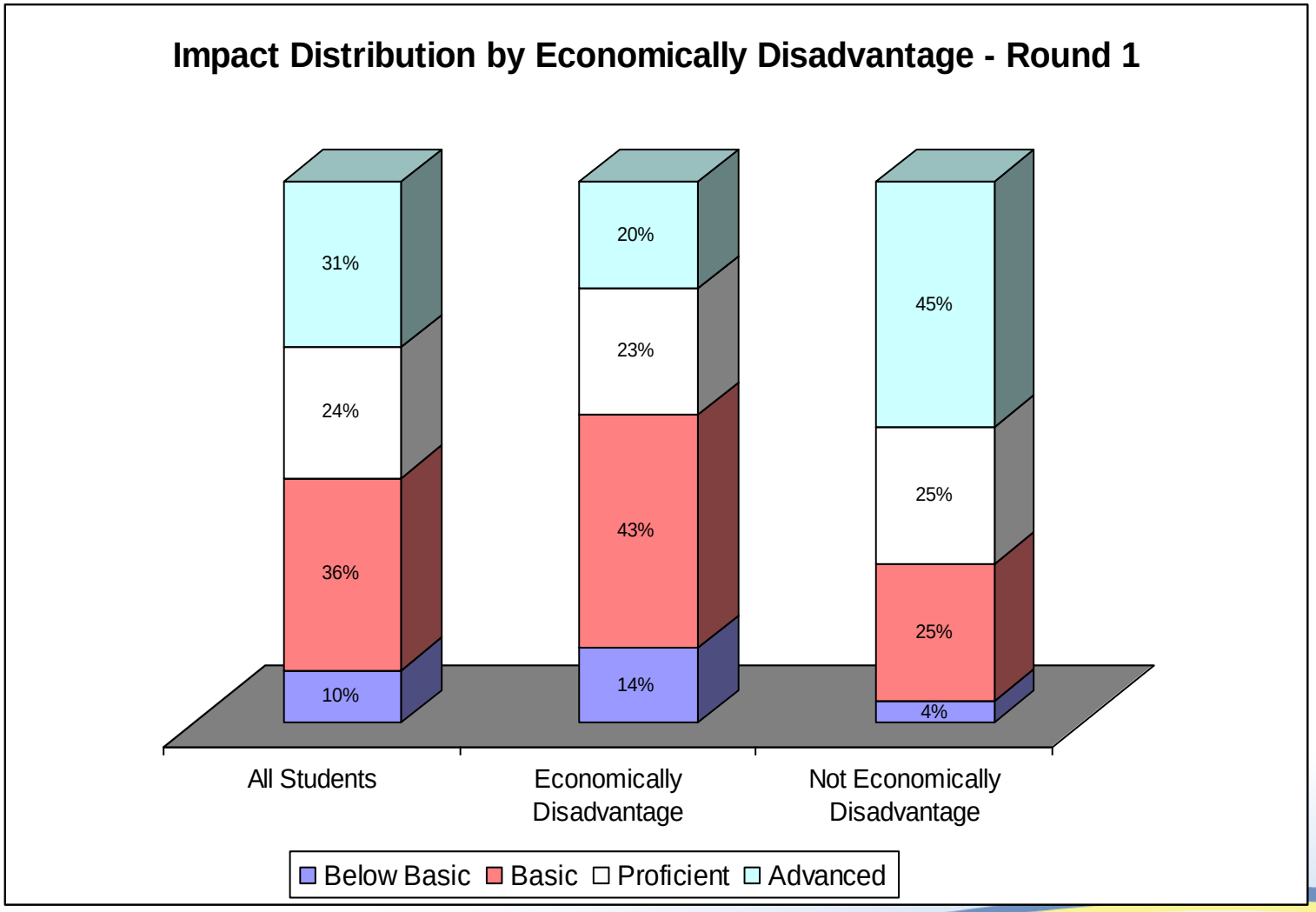
Steps in the Bookmark Process

7. Provide feedback on Round 1 recommendations
   - Summary of bookmarks for each table and total group
   - P-values
   - Impact associated with median bookmark recommendations

Mathematics, Grade 3, Round 1 Table 1

	Basic	Proficient	Advanced
Minimum	9	39	60
Maximum	14	47	60
Mean	1	43	60
Median	**9**	**45**	**60**

Impact Data Feedback



Standard Setting - Process

8. Make Round 2 recommendations (lower grade)
9. Review ALDs/Define threshold students (upper grade)
10. Make Round 1 and 2 recommendations
11. Review and discuss Round 2 results across grades
12. Make final round of recommendations
    - provide feedback to support vertical articulation

Vertical Articulation

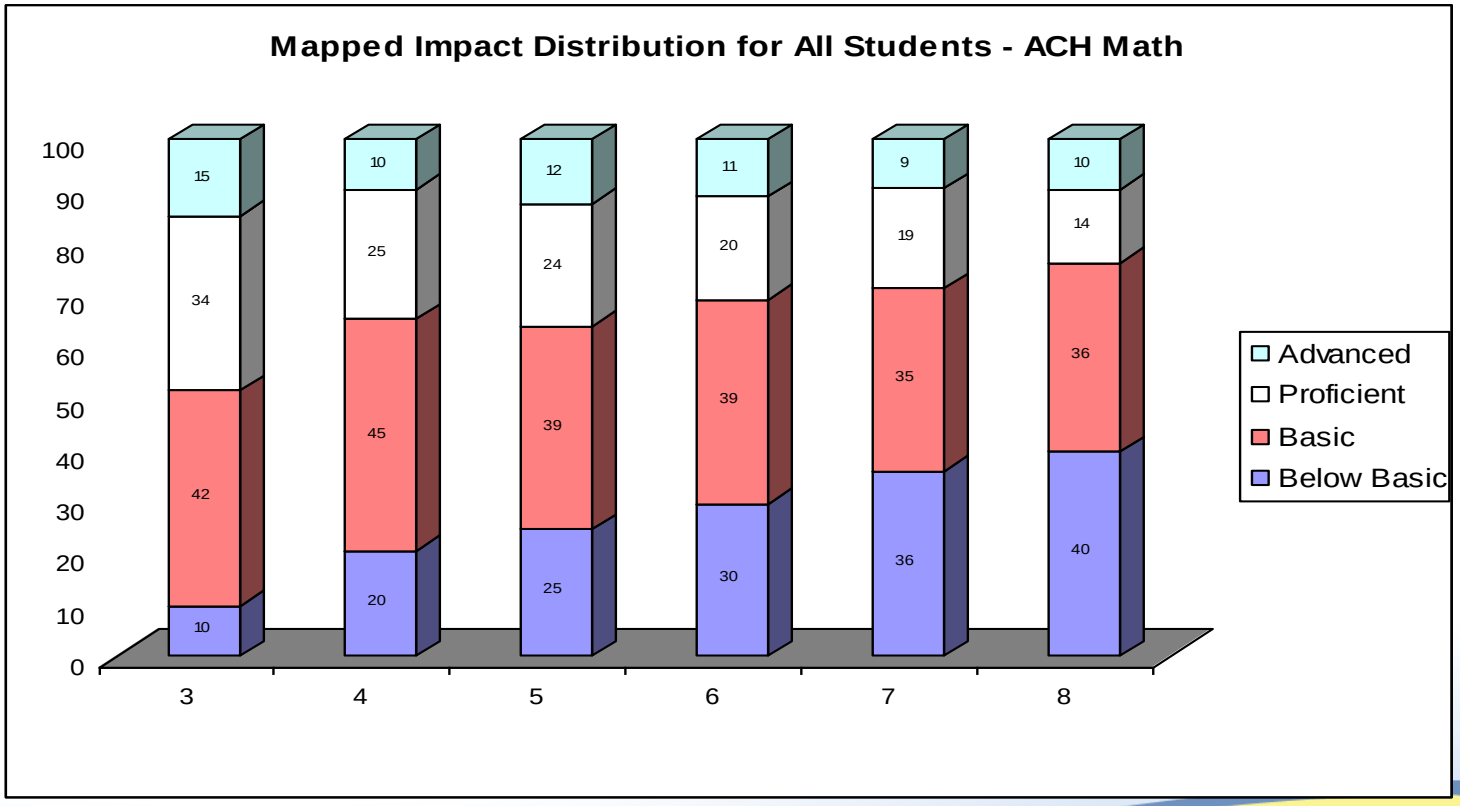
- Participants:
  - 1/3 of panelists from each committee
- Purpose:
  - Review impact within and across grades for a given content area to see if it makes sense given: ALDs, Test taking populations, skills assessed
- Task:
  - Make a final recommendation as to what the impact should look like across grades.
    - Stay true to the content-based recommendation.
    - Consider group discussion, ALDs and expectations to mildly smooth results.

Steps in the Process

- Review ALDs for all grades
- Discuss impact expectations across grades
- Review impact associated with Round 3 median bookmark recommendations
- Review summary of Rd 2 impact feedback
- Panelist make independent impact recommendations
- Review min., max., median and mean impact recommendation over all panelists
- Finalize impact recommendation
- Map recommendations back to the observed frequency distribution

End Result of Standard Setting Meeting

- Cut score recommendation associated with each test.
- Chart detailing impact associated with those recommendations for 2010.



Questions?

Erika.Hall@Pearson.com